

ABSTRACT

A system and method of generating morphemes from received speech is disclosed. The method comprises selecting candidate sub-morphemes from the received speech, selecting salient sub-morphemes from the candidate sub-morphemes based on salience measurements and clustering the salient sub-morphemes based on similar characteristics into morphemes. The generated morphemes may be acoustic or non-acoustic morphemes and are used in speech recognition and speech understanding modules of natural language dialog systems.